

## **Supporting Information (SI) Text**

### **Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris***

Gavin Douglas<sup>a,1</sup>, Gesseca Gos<sup>a,1</sup>, Kim Steige<sup>b,1</sup>, Adriana Salcedo<sup>a</sup>, Karl Holm<sup>b</sup>, J. Arvid Ågren<sup>a</sup>, Khaled Hazzouri<sup>a,c</sup>, Wei Wang<sup>a</sup>, Adrian Platts<sup>d</sup>, Emily B. Josephs<sup>a</sup>, Robert J. Williamson<sup>a</sup>, Barbara Neuffer<sup>e</sup>, Martin Lascoux<sup>b,2</sup>, Tanja Slotte<sup>b,f,2</sup>, Stephen I. Wright<sup>a,2</sup>

1 These authors contributed equally to the work

2 authors for correspondence, email:

stephen.wright@utoronto.ca

tanja.slotte@su.se

martin.lascoux@ebc.uu.se

a Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

b Department of Ecology and Genetics, Evolutionary Biology Centre, Science for life Laboratory, Uppsala University, Uppsala, Sweden

c Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

d McGill Centre for Bioinformatics, McGill University, Montreal, Canada

e Department of Botany, University of Osnabrück, Osnabrück, Germany

f Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

## Genome Assembly, Variant Calling and Phasing

A fragment-level assembly of *Capsella bursa pastoris* was undertaken with the aim of generating KBase size-range sequences that would be informative for a gene-level realignment against *Capsella rubella*, rather than for a comprehensive genome assembly per se. The low sequence divergence between homeologous sequences poses challenges for assembly and argued for a relatively conservative approach, as a consequence of which the Ray assembler (1) was selected over more aggressive assemblers (eg Velvet, EBI). Approximately 20 million 2x105nt illumina read pairs (4.3GBase) from two European accessions (12.4 and 16.9) were 3' trimmed to remove sequences with a quality <30 and contigs were then generated by the Ray assembler v 1.4 (1) using a Kmer length of 31. A marginal degree of further scaffolding was then undertaken with the Soap de Novo assembler (BGI, v 1.05) using the same paired-end data and a Kmer length of 51. However this only slightly extended the Ray contigs. The final scaffold N50 was 2.5KBase (0.23% uncalled bases), contig N50 2.4KBase, with a maximum scaffold/contig length of ~25KBase. The total genome size was 210.5MBase with 83% of the genome present in sequences longer than 0.5KBase.

For polymorphism analyses, Illumina reads for both *C. orientalis* and *C. bursa-pastoris* were mapped to the *Capsella rubella* reference genome(2), using Stampy version 13 (3), and Picard (picard.sourceforge.net) was used for read sorting and file format conversion. Genotyping of SNPs and short indels was done using the Genome Analysis Toolkit (GATK) software package (4, 5), with re-alignment of

sequences surrounding short indels (6). The resulting polymorphism data was combined with similarly processed whole genome polymorphism data from 13 *C. grandiflora* individuals (7, 8).

To minimize the risk of spurious SNP calling, each site in the VCF files was rigorously filtered by PHRED quality score of 40 and depth. *Capsella bursa-pastoris* and *C. grandiflora* VCF file depth cut-offs were a minimum of 20 and maximum of 100. Depth cut-offs for *C. orientalis* were a minimum of 15 and maximum of 100. We also excluded 20kb genomic windows where less than 30% of sites passed these quality and depth criteria for a given species. Many of these regions corresponded to pericentromeric regions of the genome.

To identify transposable element insertions in *C. bursa-pastoris* we used the Popoolation TE approach developed by Kofler et al. (9). Using default settings, we ran the pipeline on paired-end Illumina (108 bp) samples from 8 of our *C. bursa-pastoris* individuals, using the *C. rubella* genome (2) as the reference genome. Since Popoolation TE was originally developed for pooled population samples designed to infer population wide frequencies, we modified it to apply to individual samples using frequency cutoffs to define heterozygous and homozygous insertions (J.A. Ågren, W. Wang, D. Koenig, B. Neuffer, D. Weigel and S.I. Wright, in revision). Principal components analysis of insertion presence/absence in *C. bursa-pastoris* was analyzed along with insertions previously identified in our samples of *C. orientalis* and *C. grandiflora* (J.A. Ågren, W. Wang, D. Koenig, B. Neuffer, D. Weigel and S.I. Wright, in revision) using SPSS v. 22 (SPSS Inc. 2009) under default settings.

Phasing of the *C. bursa-pastoris* homeologs was conducted with HapCUT v0.6 (10), which has been used to successfully phase tetraploid wheat (11). HapCUT was run on SNPs passing our stringent heterozygosity and depth filters, and the resulting phased blocks were thus composed of high-confidence polymorphisms. Each *C. bursa-pastoris* sample was phased individually and then phased blocks were identified as either unknown origin, descended from *C. grandiflora* or *C. orientalis* or discordant between the two based on SNPs shared with the progenitors. SNPs identified within these blocks were compared across all samples and any block containing a SNP inconsistently assigned to different homeologs was removed from downstream analyses. This procedure could lead us to globally underestimate polymorphism levels in *C. bursa-pastoris*, but should not bias our inference of selection or affect inference of biased fractionation, as no particular category of SNPs is preferentially removed, and both homeologs are expected to be equally affected by this procedure. In total 523,425 SNPs were used to infer the parental origin of the phased blocks with 16% of these resulting in discordant assignments across samples. The procedure was validated by comparison to Sanger sequence data for six independently amplified and sequenced genomic fragments from the two homeologs. Principal Components analysis on phased SNP was run using the resulting dataset, with analysis restricted to common SNPs at a frequency of 6 or more across the entire dataset. PCA of the phased SNPs was run using SPSS v. 22 (SPSS Inc. 2009) under default settings

### *Comparative Genomics Analyses*

A multiple whole-genome alignment between the *C. rubella* reference genome (2) and Illumina fragment assemblies of *C. bursa-pastoris*, *C. grandiflora* (2), *C. orientalis* (using Ray v. 1.4, J.A. Ågren, W. Wang, D. Koenig, B. Neuffer, D. Weigel and S.I. Wright, in revision) and *Neslia paniculata* (2) was conducted essentially as described by Haudry and colleagues (7). Briefly, each fragment-assembly was initially aligned against the soft masked (RepeatMasker, [www.repeatmasker.org](http://www.repeatmasker.org)) *C. rubella* reference sequence using Lastz (12) with parameters --gapped --nochain --gfextend --strand=both. Alignments were then chained using axtChain (Kent tools, UCSC) with a minimum chain Score of 10,000 and a slightly customized linear gap table. Chains were then selected for the subset of most likely orthologous chains having the maximum alignment score against the *C. rubella* reference, retaining only a single chain for each *C. rubella* sequence for all assemblies except *C. bursa-pastoris* in which up to two chains could be selected providing there was sufficient evidence for two good orthologous alignments. Finally the individual alignments against *C. rubella* were iteratively merged by phylogenetic distance using multiz (13) with default parameters to create a multiple alignment.

We constructed maximum-likelihood phylogenies for alignments of each assembled fragment that included two distinct *C. bursa-pastoris* sequences as well as the orthologous *C. grandiflora*, *C. rubella*, *C. grandiflora*, and *Neslia paniculata* sequences. We constructed phylogenies using RAxML's (14) rapid bootstrap algorithm to find the best-scoring ML tree. Each phylogeny had 100 bootstrap replicates and used *N. paniculata* as the outgroup. We excluded trees with less than 80% bootstrap support at any branch from further analysis. We then used a custom

Perl script to count the number of resulting phylogenies corresponding to each topology.

We used two approaches to validate our phylogenetic inference. First, we assessed whether similar patterns were observed with Sanger data and larger sample sizes for *C. bursa-pastoris*. Second, we assessed whether patterns of fixed heterozygosity and fixed differences between the diploid putative ancestors were in agreement with expectations under our phylogenetic hypothesis.

For the first validation, we assessed phylogenetic patterns at nine independent nuclear genes, where both homeologs have previously been amplified and sequenced in *C. bursa-pastoris* with homeolog-specific primers (15-17). We amplified and sequenced the same loci in *C. orientalis* using these previous primers and, and used Muscle (18) to align our sequences to publicly available data for the same loci from *C. grandiflora* (sequences phased using PHASE2.1; (19, 20), *C. rubella* and *C. bursa-pastoris* from (16, 17, 21-25) (Table S4). As outgroups, we used *Arabidopsis thaliana* and/or *N. paniculata*. All positions with gaps or missing data were removed and data for each locus was collapsed into unique haplotypes using FaBox 1.40 (<http://users-birc.au.dk/biopv/php/fabox/>). Subsequently neighbor-joining trees were reconstructed in MEGA5 (26), with distances estimated based on the maximum composite likelihood method (27) and support evaluated using 1000 bootstrap replicates.

### *Population Genetic Analyses*

#### Demographic inference

To infer demographic parameters associated with allopolyploid speciation in *C. bursa-pastoris*, we analyzed site frequency spectra for 60,225 SNPs at intergenic non-conserved regions and four-fold synonymous sites. Specifically, we used fastsimcoal2.1 (28) to infer demographic parameters based on the multidimensional site frequency spectrum for *C. grandiflora*, *C. orientalis*, and the two *C. bursa-pastoris* homeologous genomes. Estimates were obtained using the composite maximum likelihood approach, under four models that differed in the type of population size change (stepwise or exponential) allowed, and in the presence or absence of post-polyploidization asymmetric migration (Figure 3). All parameter estimates were global maximum likelihood estimates from 50 independent fastsimcoal2.1 runs, with a minimum of 50,000 and a maximum of 250,000 coalescent simulations, and 10-40 cycles of the likelihood maximization algorithm. Multidimensional SFS entries with less than 5 SNPs were pooled to avoid negative effects on the estimation procedure, as suggested by Excoffier and colleagues (28). We assumed a mutation rate of  $7 \times 10^{-9}$  (29) and a generation time of 1 year when converting estimates to units of years and individuals. Confidence intervals of parameter estimates were obtained by parametric bootstrapping, with 100 bootstrap replicates per model. Model fit was assessed using the Akaike information criterion and Akaike's weight of evidence, as in (28). Note that, because of possible linkage disequilibrium among our SNPs, particularly in selfing populations, confidence intervals and the strength of AIC model support should be treated with some caution. However the parameter estimates themselves under the composite likelihood approach are expected to be robust, and we obtained comparable timing estimates across models (Figure 4) and

when using different subsets of sites with likely differences in their LD structure (nonconserved noncoding sequence and 4-fold degenerate sites), arguing that our main conclusions are not affected by LD between SNPs. Furthermore, to further investigate the possible influence of linkage disequilibrium on the demographic inference we re-ran the models excluding sites less than 10kb apart, and the major conclusions, including the best-fitting model, were found to be unaffected

#### Estimating the distribution of fitness effects

Estimation of the distribution of deleterious fitness effects (DFE) was calculated using the maximum likelihood approach designed by Keightley and Eyre-Walker (30), which uses the allele frequency spectrum of polymorphic sites to infer the strength of purifying selection (see estimates of polymorphism above).

The subset of sites considered were those that passed filtering criteria within all three *Capsella* species. Frequency data was taken from *C. bursa-pastoris*, *C. grandiflora*, and *C. orientalis* for several site classes: 0-fold non-synonymous, 4-fold synonymous, and conserved noncoding sites (CNCs) that were identified in a comparison of nine Brassicaceae species using the *Capsella* reference genome as the reference for alignment (8). Sites that were heterozygous in all individuals of *C. bursa-pastoris* (hereafter referred to as ‘fixed heterozygotes’) were not included in these analyses, as they are not segregating within either homeolog. To ensure an accurate comparison between species, the *C. grandiflora* dataset was downsampled to 10 individuals. Heterozygous sites in *C. grandiflora* were converted to homozygotes by randomly selecting one of the two bases, in order to mitigate any

bias caused by higher heterozygosity in this species. To determine 95% confidence intervals, 200 bootstrap replicates of each site class were used to recalculate the folded allele frequencies and numbers of divergent sites. Significance was determined as described in Keightley & Eyre-Walker (30). To account for uncertainty due to linkage disequilibrium when estimating the DFE, significance tests and bootstrap confidence intervals were generated by resampling 10kb blocks of SNPs.

#### *Inference of Selective Effects and Functional Enrichment*

##### Inference of gene loss

We took two distinct approaches to identify putative deletion events. First, we used HTseq (31) to count the number of reads mapping to each annotated gene in the *Capsella* reference genome, using the ‘intersection\_nonempty’ option. After normalizing each sample by the total number of reads, we identified genes that showed significant reductions in coverage in the *C. bursa pastoris* samples in paired tests of both *C. grandiflora* and *C. orientalis*. Significance was assessed using 2-tailed t tests assuming unequal variance. Only tests with *p* values less than 0.01 against both species (corresponding to a false discovery rate FDR of approximately 5%) were treated as significant. Furthermore, to restrict our analyses to putative single-copy deletion events rather than variance in read mapping success and/or high copy number genes, we only considered cases where the fold change in coverage in *C. bursa-pastoris* ranged from 0.25 to 0.65.

As a second approach, we used Pindel (32) to identify large deletions spanning whole genes and small insertion/deletion events affecting coding regions. Pindel was run for each sample independently, and compared to the gene annotation. Gene deletions were called as deletions covering 80% of a locus. Overlapping variants between individuals and species were identified with the bedtools intersect command. Gene deletions and inversions required 80% overlap to be called as orthologous, shorter indels required complete overlap.

#### Effect Prediction of Polymorphisms

The software package SnpEff v3.5 was used to predict the genomic effects of SNPs and structural variants, given the *Capsella rubella* genome annotation (33). Mutations more likely to cause loss-of-function effects were identified by using the option “-lof” and parsing for mutations flagged for “HIGH” effect. This set included polymorphisms knocking out splice sites, start or stop codons, or causing the gain of stop codons as well frameshift deletions and short-insertions. Possible compensatory mutations for all of these putatively deleterious mutations were accounted for within 50bp of each mutation. For instance, frameshift mutations within the same gene and within 50bp of each other were excluded from the analysis. Polymorphisms fixed between *C. bursa-pastoris* homeologs (segregating at exactly 50% frequency) were included as a separate category, in order to test for the fixation of deleterious mutations within homeologs. Since all mutations were called relative to the *C. rubella* assembly, all mutations were polarized by using one *Neslia*

*paniculata* individual as an outgroup. Any putatively deleterious mutation also found in *N. paniculata* was excluded from the analysis.

### Functional Category Enrichment

Gene ontologies (GO), the classification of genes into classes of molecular function, cellular components and biological process, were inferred from *Arabidopsis thaliana*, using the Virtual Plant online server (version 1.3, (34)). Since *Capsella* is closely related to *Arabidopsis*, the majority of orthologous genes in *A. thaliana* are likely to have the same function in the *Capsella* species. A total of 19,520 genes in *Arabidopsis* that have known orthologs in *C. rubella* were included in this analysis.

Genes containing putatively deleterious SNPs, according to the SnpEff algorithm, in *C. bursa-pastoris* were analyzed with Virtual Plant's BioMaps tools (34). Single nucleotide polymorphisms fixed on a single homeolog ("fixed heterozygotes") were considered separately. Genes associated with GO categories that were found to be enriched among singletons retained following the ancient WGD event in *Arabidopsis* (35) were acquired from the Virtual Plant database. Only GO categories associated with less than one thousand genes were included for downstream analyses. Categories including *organelle* (6778 genes), *cell part* (14508 genes) and *cellular metabolic process* (5998 genes) are examples of excluded groups. These GO terms are parents of the smaller GO classes included, so their removal likely reduced noise without compromising statistical power.

## Scanning for Selection

Polymorphisms at 4-fold synonymous sites were counted in 1000, 10 000 and 50 000 base-pair windows in coding regions upstream and downstream of selected putatively deleterious mutations in *C. bursa-pastoris*, and windows were averaged over mutations of the same category. The numbers of these neutral mutations was normalized by divergence to *N. paniculata*. Normalizing by divergence controls for differences in mutation rate across the genome. The SNPs fixed on a single homeolog were considered separately. Singleton genes whose orthologs were retained following the ancient genome duplication in *Arabidopsis* were scanned (35), as were related genes from the same GO categories. Finally, nuclear genes associated with the chloroplast were considered separately, since they were significantly enriched for loss-of-function mutations in past analyses. VCFtools v0.1.12a was also used to assess nucleotide diversity over all site types in the same size windows surrounding fixed putatively deleterious mutations using the “--window-pi” option. Neutral expectations of diversity surrounding a given fixation were generated by analyzing diversity in 50kb windows surrounding 4-fold synonymous fixations. 95% CI for each window were computed through bootstrapping by substitution (n=1000). The result show in Figure S5 is robust to all window sizes, although noisier at smaller window sizes.

1. Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17:1519–1533.
2. Slotte T et al. (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45:831–U165.

3. Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* 21:936–939.
4. McKenna A et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
5. DePristo MA et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
6. Van der Auwera GA et al. (2002) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in ...*:11.10.1–11.10.33.
7. Haudry A et al. (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet* 45:891–U228.
8. Williamson RJ et al. (2014) *Evidence for widespread positive and negative selection in coding and conserved noncoding regions of Capsella grandiflora*.
9. Kofler R, Betancourt AJ, Schlötterer C (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet* 8:e1002487.
10. Bansal V, Bafna V (2008) HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. *Bioinformatics* 24:i153–9.
11. Krasileva KV et al. (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol* 14:R66.
12. Harris RS (2007) Improved Pairwise Alignment of Genomic DNA. - Robert S. Harris - Google Books.
13. Blanchette M et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research* 14:708–715.
14. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
15. Slotte T, Ceplitis A, Neuffer B, HURKA H, Lascoux M (2006) Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *Am J Bot* 93:1714–1724.
16. Slotte T, Huang H, Lascoux M, Ceplitis A (2008) Polyploid speciation did not

confer instant reproductive isolation in *Capsella* (Brassicaceae). *Mol Biol Evol* 25:1472–1481.

17. Slotte T et al. (2009) Splicing variation at a FLOWERING LOCUS C homeolog is associated with flowering time variation in the tetraploid *Capsella bursa-pastoris*. *Genetics* 183:337–345.
18. Edgar RC (2004) BMC Bioinformatics | Full text | MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*.
19. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169.
20. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.
21. Guo Y-L et al. (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci USA* 106:5246–5251.
22. Slotte T, Foxe JP, Hazzouri KM, Wright SI (2010) Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol Biol Evol* 27:1813–1821.
23. Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D (2011) Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biol Evol* 3:868–880.
24. St Onge KR et al. (2012) Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd's Purse (*Capsella bursa-pastoris*). *Mol Biol Evol* 29:1721–1733.
25. St Onge KR, Källman T, Slotte T, Lascoux M, Palmé AE (2011) Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. *Mol Ecol* 20:3306–3320.
26. Tamura K et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739.
27. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 101:11030–11035.

28. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC (2013) PLOS Genetics: Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.*
29. Ossowski S, Schneeberger K, Lucas-Lledó JI (2010) The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science.*
30. Keightley PD, eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177:2251–2261.
31. Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. *bioRxiv.*
32. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25:2865–2871.
33. Cingolani P et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.
34. Katari MS et al. (2010) VirtualPlant: A Software Platform to Support Systems Biology Research. *Plant Physiol* 152:500–515.
35. De Smet R et al. (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci USA* 110:2898–2903.

Figure S1

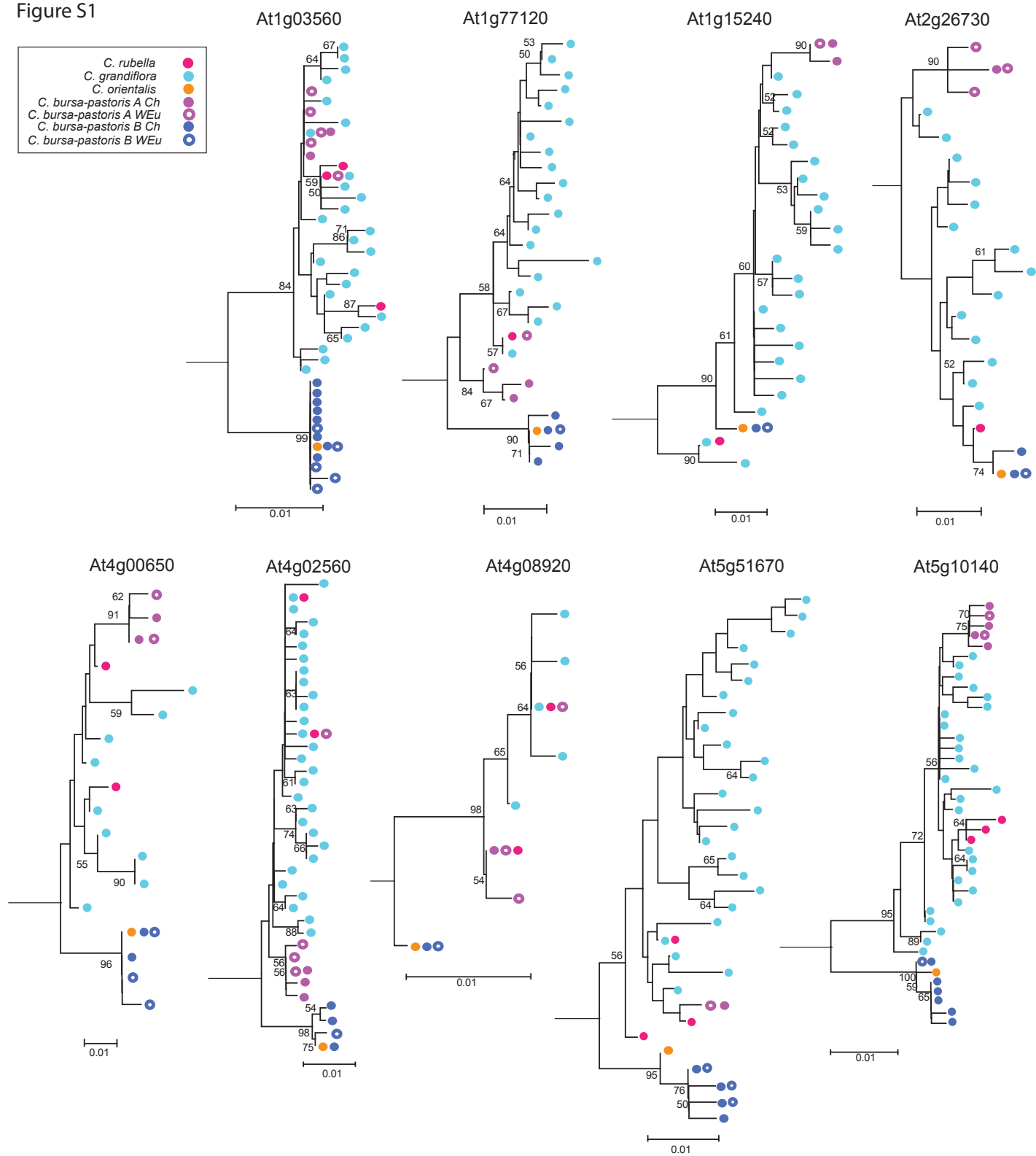


Figure S2

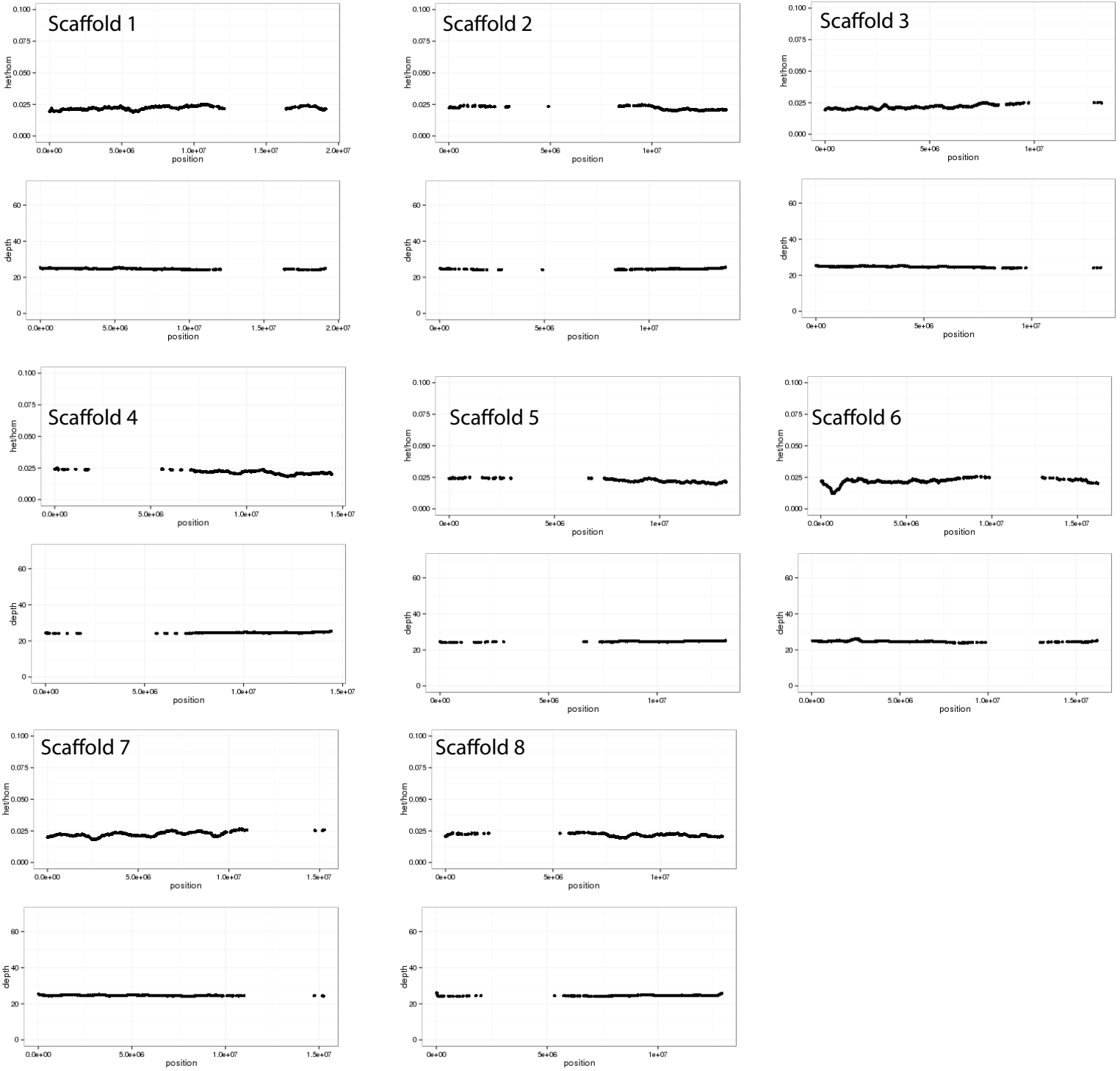


Figure S3

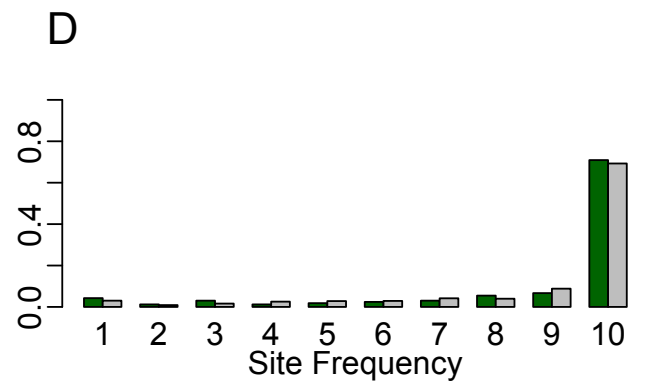
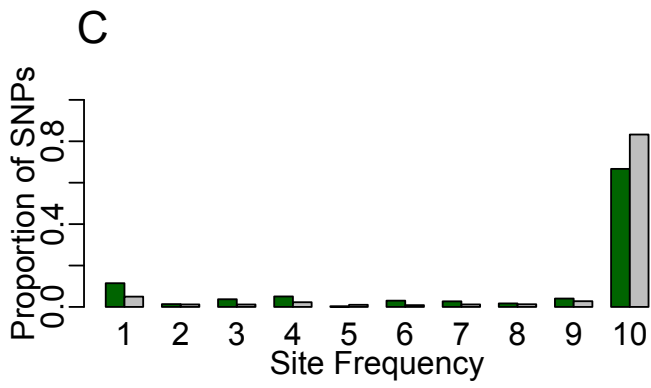
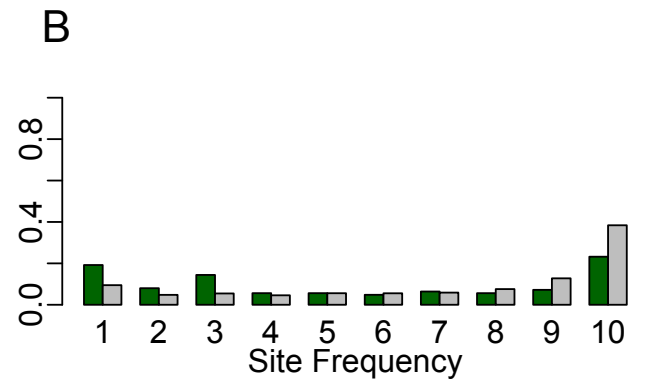
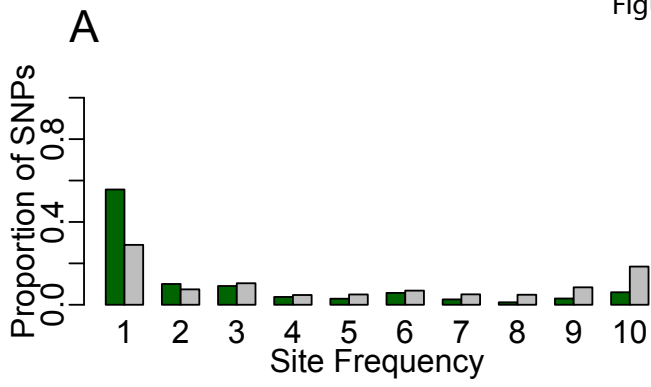


Figure S4

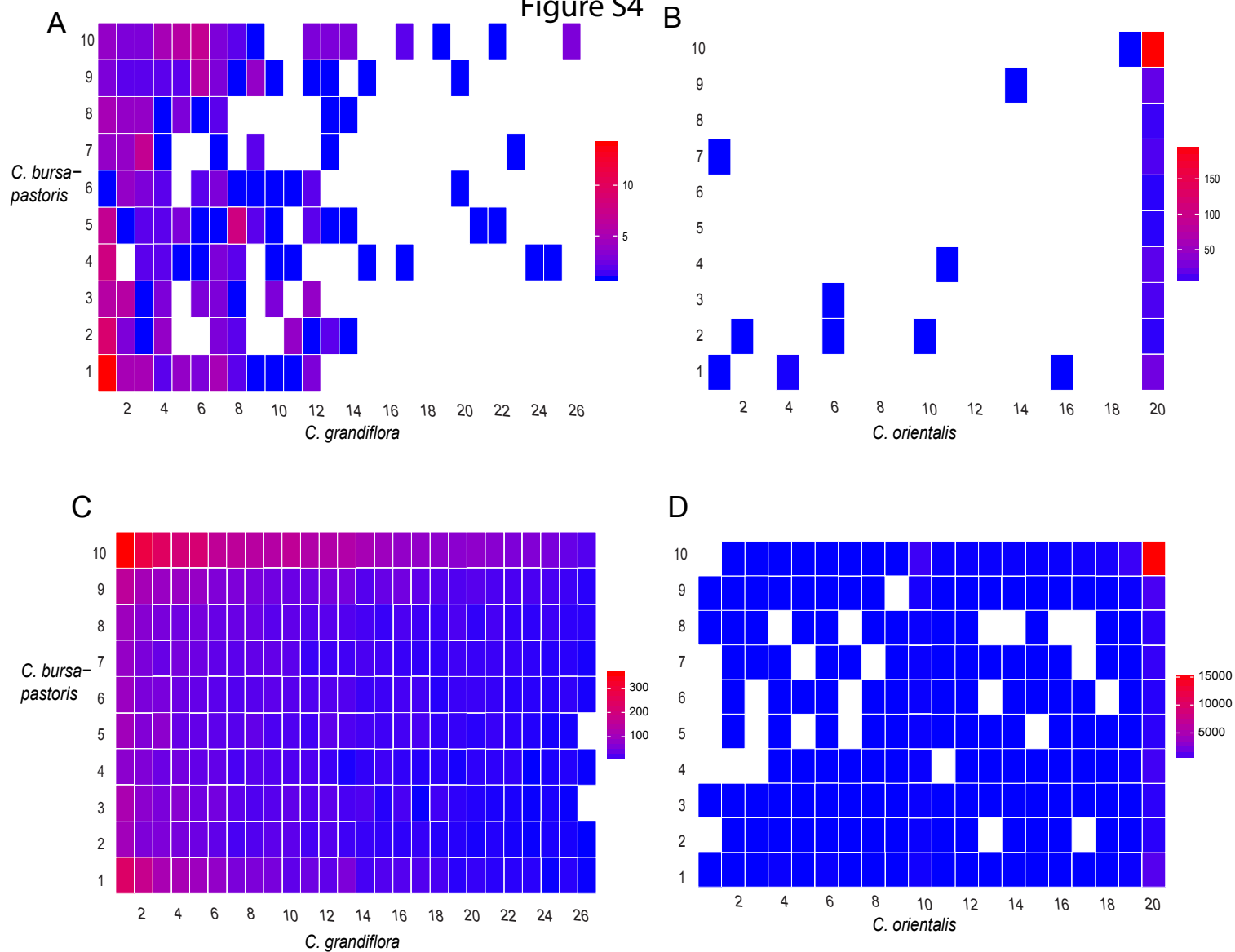
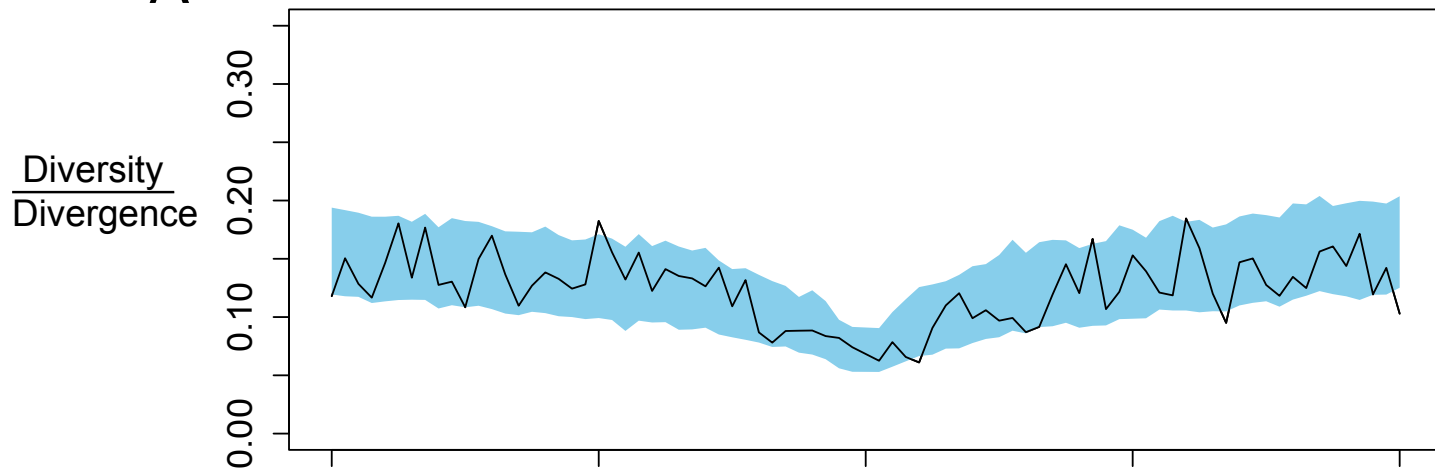


Figure S5

A



B

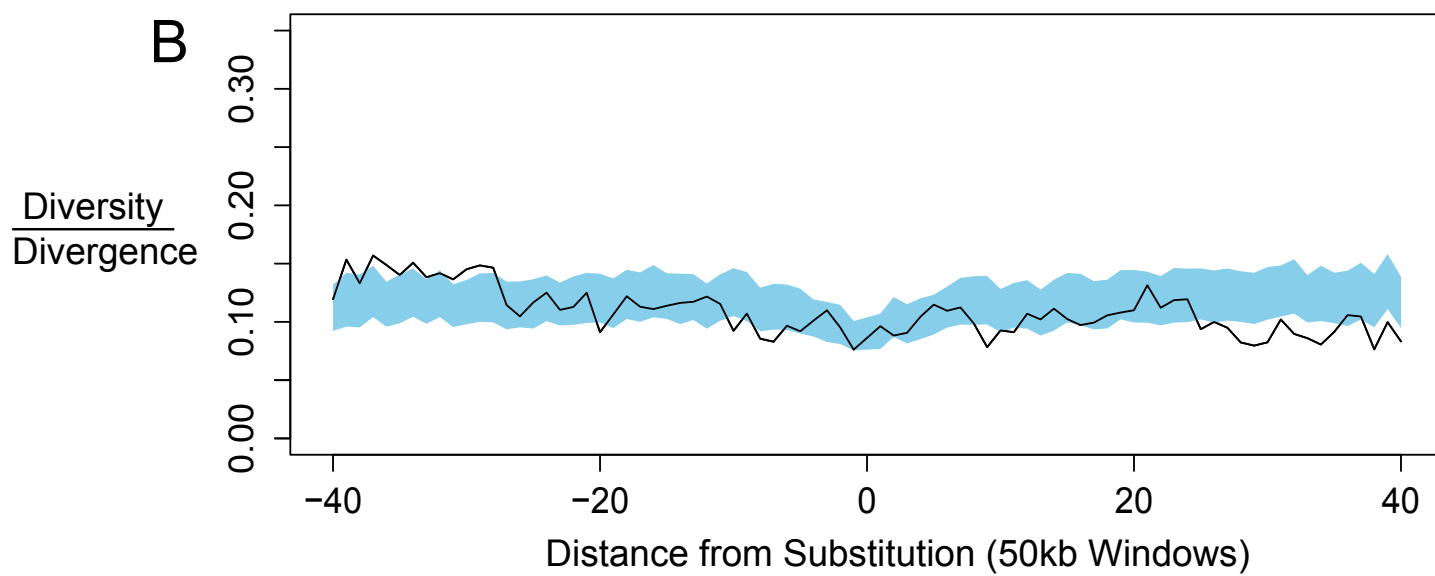
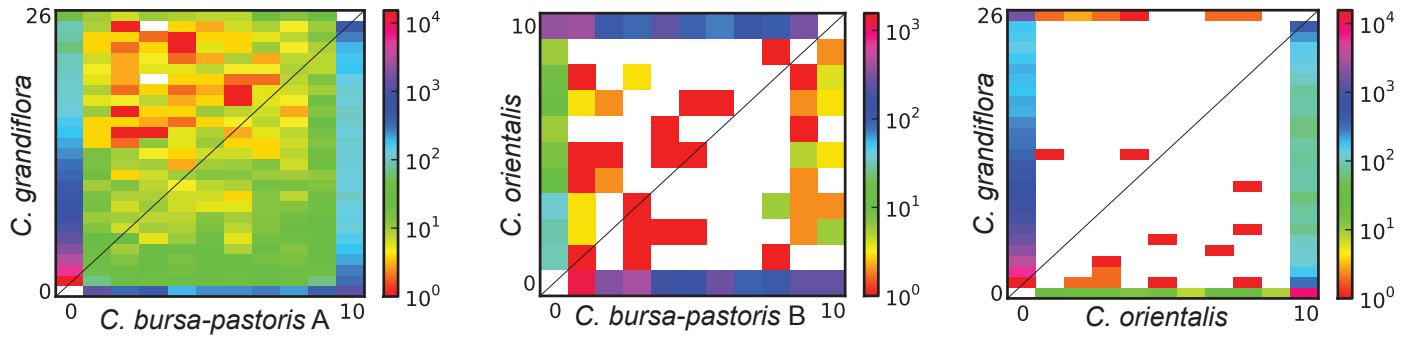
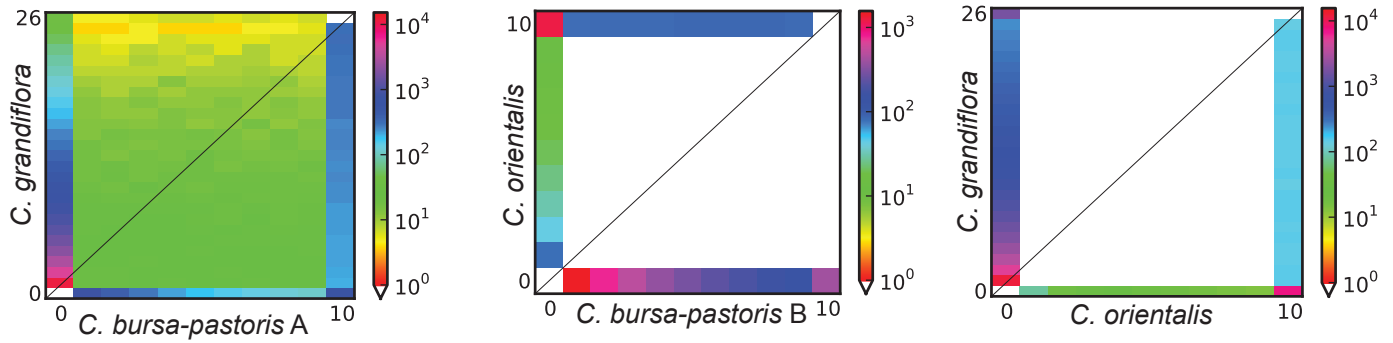


Figure S6

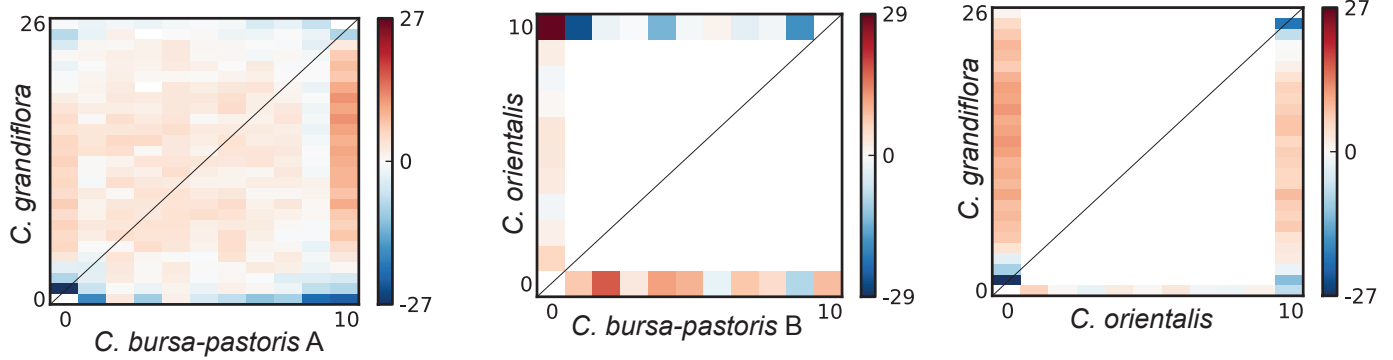
A. data



B. model



C. residuals



D. residuals

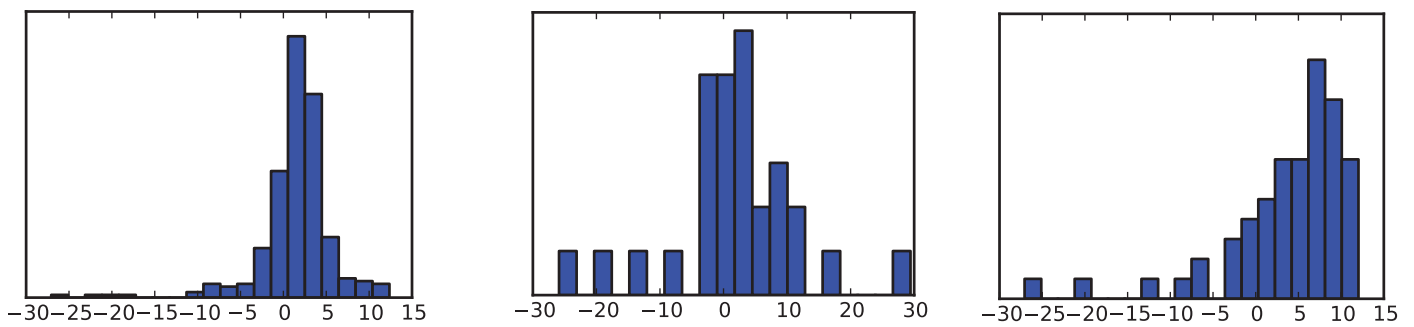


Table S1-Gene loss events identified using coverage analysis in C. bursa-pastoris

| gene                    | centromeric region? | C. rubella scaffold | gene start position | average fold change in Cbp | clear breakpoint in assembly? | Arabidopsis UCSC orthologue? | Closest BLAST Arabidopsis hit in the absence of UCSC browser orthologue | C. bursa-pastoris scaffold | breakpoint 1 | breakpoint 2 | length   | sign of clustered deletion? | type of gene                                                                    |                                         |
|-------------------------|---------------------|---------------------|---------------------|----------------------------|-------------------------------|------------------------------|-------------------------------------------------------------------------|----------------------------|--------------|--------------|----------|-----------------------------|---------------------------------------------------------------------------------|-----------------------------------------|
| PAC-2088858             |                     |                     | 1                   | 4312843                    | 0.577477088 no                | no                           | AT3G27999.1                                                             |                            |              |              |          | no                          | invertase                                                                       |                                         |
| PAC-20888727            |                     |                     | 1                   | 4713681                    | 0.56 yes                      | no                           | AT1G12855.1                                                             | scaffold10630              | 4713568      | 4714635      | 1067     | no                          | F-box protein                                                                   |                                         |
| PAC-20893036            |                     |                     | 1                   | 8340736                    | 0.340470757 yes               | AT1G23660                    |                                                                         |                            |              |              |          | yes                         | unknown                                                                         |                                         |
| PAC-20891992            |                     |                     | 1                   | 8342484                    | 0.393613652 yes               | AT1G23670                    |                                                                         |                            |              |              |          | yes                         | unknown                                                                         |                                         |
| PAC-20890788            |                     |                     | 1                   | 8344580                    | 0.435928066 yes               | no                           | AT1G23600                                                               | scaffoldc282325            | 8342214      | 8349370      | 7156     | yes                         | unknown                                                                         |                                         |
| PAC-20888826            |                     |                     | 1                   | 8469794                    | 0.49965 yes                   | no                           | AT1G24417                                                               | c326973                    | 8469874      | 8471212      | 1338     | no                          | unknown                                                                         |                                         |
| PAC-20893281            |                     |                     | 1                   | 9938281                    | 0.57 yes                      | AT1G29110                    |                                                                         | c302661                    | 9938424      | 9939661      | 1237     | no                          | cysteine peptidase                                                              |                                         |
| PAC-2089255 centromeric |                     |                     | 1                   | 14752839                   | 0.522484447 no                |                              | HEUTRON                                                                 |                            |              |              |          | yes                         | HEUTRON-TE                                                                      |                                         |
| PAC-2089151 centromeric |                     |                     | 1                   | 14759628                   | 0.298549722 no                |                              | At2g13542                                                               |                            |              |              |          | yes                         | repetitive, possible defensin-like protein                                      |                                         |
| PAC-2089255 centromeric |                     |                     | 1                   | 16577225                   | 0.609 yes                     | AT1G78470                    |                                                                         | scaffold18523              | 16576712     | 16578034     | 1322     | no                          | F-box protein                                                                   |                                         |
| PAC-2089111 centromeric |                     |                     | 1                   | 17573455                   | 0.5016 no                     | AT1G50800                    |                                                                         |                            |              |              |          | no                          | plant mobile domain protein family                                              |                                         |
| PAC-20906370            |                     |                     | 2                   | 2357403                    | 0.574935973 no                | no                           | AT1G59780.1                                                             |                            |              |              |          | no                          | duplicated in reference, NB-ARC domain-containing disease resistance protein    |                                         |
| PAC-2090681 centromeric |                     |                     | 2                   | 3533871                    | 0.389619892 no                | no                           | AT1G18250.2                                                             |                            |              |              |          | no                          | thaumatin superfamily-pathogenesis?                                             |                                         |
| PAC-2090602 centromeric |                     |                     | 2                   | 4114216                    | 0.56 no                       | AT1G54890                    |                                                                         |                            |              |              |          | no                          | late embryogenesis abundant protein                                             |                                         |
| PAC-2090641 centromeric |                     |                     | 2                   | 8038731                    | 0.487273276 no                | no                           |                                                                         |                            |              |              |          | no                          |                                                                                 |                                         |
| PAC-2090445 centromeric |                     |                     | 2                   | 8081418                    | 0.548276439 no                | no                           | AT1G15870                                                               |                            |              |              |          | no                          | terpenoid synthase                                                              |                                         |
| PAC-2090544 centromeric |                     |                     | 2                   | 8781063                    | 0.554271951 yes               | no                           |                                                                         | scaffold28450              | 8781181      | 8785301      | 4120     | no                          | GDSL-like lipase/Acylhydrolase superfamily protein                              |                                         |
| PAC-20905727            |                     |                     | 2                   | 10262025                   | 0.567007146 no                | no                           | AT1G71320                                                               |                            |              |              |          | no                          | F-box                                                                           |                                         |
| PAC-20905161            |                     |                     | 2                   | 10688866                   | 0.488383039 no                | no                           | AT1G72110                                                               |                            |              |              |          | yes                         | O-acetyltransferase protein, tandem repeat region                               |                                         |
| PAC-20906937            |                     |                     | 2                   | 10693845                   | 0.400371558 no                | no                           | AT1G72110                                                               |                            |              |              |          | yes                         | O-acetyltransferase protein, tandem repeat region                               |                                         |
| PAC-20906636            |                     |                     | 2                   | 12118192                   | 0.535 no                      | no                           | AT5G4420                                                                |                            |              |              |          | no                          | defensin 1 protein; tandem repeat region                                        |                                         |
| PAC-20898472            |                     |                     | 3                   | 4062729                    | 0.611895322 no                | AT3G12510                    |                                                                         |                            |              |              |          | no                          | MADS box transcription factor                                                   |                                         |
| PAC-20897409            |                     |                     | 3                   | 7819573                    | 0.534014472 yes               | no                           | AT3G18910                                                               | scaffold51370              | 7819154      | 7821506      | 2352     | no                          | F-box, EIN-targeting protein                                                    |                                         |
| PAC-20898312            |                     |                     | 3                   | 8757307                    | 0.545530334 no                | no                           | AT2G13295                                                               |                            |              |              |          | no                          | protease inhibitor/seed storage protein                                         |                                         |
| PAC-20900818            |                     |                     | 3                   | 8944635                    | 0.501982345 yes               | no                           | HEUTRON                                                                 | scaffold22475              | 8943871      | 8951789      | 7918     | no                          | nudix hydrolase homolog 5                                                       |                                         |
| PAC-20900215            |                     |                     | 3                   | 9756019                    | 0.58049715 no                 | AT2G04430                    |                                                                         |                            |              |              |          | no                          | leucine-rich repeat                                                             |                                         |
| PAC-2089801 centromeric |                     |                     | 3                   | 10341310                   | 0.624411304 no                | AT2G06040                    |                                                                         |                            |              |              |          | no                          | zinc finger                                                                     |                                         |
| PAC-2090007 centromeric |                     |                     | 3                   | 10909549                   | 0.484403821 no                | AT2G06541.1                  |                                                                         |                            |              |              |          | no                          | predicted peptide family similarity to tobacco rapid alkalization factor (RALF) |                                         |
| PAC-2089761 centromeric |                     |                     | 3                   | 14172993                   | 0.4939367651 no               | no                           | AT2G19030                                                               |                            |              |              |          | no                          | plant thionin protein family                                                    |                                         |
| PAC-2089911 centromeric |                     |                     | 3                   | 14881389                   | 0.560657301 no                | AT2G20619                    |                                                                         |                            |              |              |          | yes                         | self-incompatibility family protein                                             |                                         |
| PAC-2089851 centromeric |                     |                     | 3                   | 14929377                   | 0.568239573 no                | no                           | AT5G04047.1                                                             |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20898417            |                     |                     | 3                   | 15002693                   | 0.360951122 no                | no                           | none                                                                    |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20901125            |                     |                     | 4                   | 459563                     | 0.420599499 no                | AT1G77525                    |                                                                         |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20902735            |                     |                     | 4                   | 470895                     | 0.424878708 no                | no                           | AT2G22145.1                                                             |                            |              |              |          | no                          | ECA1 gametogenesis related family protein                                       |                                         |
| PAC-20903666            |                     |                     | 4                   | 1165440                    | 0.424072202 no                | no                           | AT2G23510                                                               |                            |              |              |          | no                          | spermidine transferase                                                          |                                         |
| PAC-2090411 centromeric |                     |                     | 4                   | 3256983                    | 0.340828282 no                | no                           | AT2G26320.1                                                             |                            |              |              |          | no                          | agamous-like 33                                                                 |                                         |
| PAC-2090311 centromeric |                     |                     | 4                   | 3787537                    | 0.596665061 no                | no                           | AT1G72950.1                                                             |                            |              |              |          | no                          | tr_NBS disease resistance gene                                                  |                                         |
| PAC-20903132            |                     |                     | 4                   | 10367196                   | 0.406962248 no                | no                           | AT2G36700.1                                                             |                            |              |              |          | no                          | pectin lyase-like superfamily                                                   |                                         |
| PAC-20902483            |                     |                     | 4                   | 11271182                   | 0.519343505 no                | no                           | AT3G27900.1                                                             |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20903193            |                     |                     | 4                   | 12381350                   | 0.581756029 no                | AT2G41342                    |                                                                         |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20885989            |                     |                     | 5                   | 514114                     | 0.537525214 no                | AT2G19802                    |                                                                         |                            |              |              |          | no                          | unknown                                                                         |                                         |
| PAC-20886869            |                     |                     | 5                   | 912215                     | 0.487822984 no                | no                           | AT2G03030                                                               |                            |              |              |          | no                          | TR domain defence gene                                                          |                                         |
| PAC-2088721 centromeric |                     |                     | 5                   | 3319219                    | 0.602084988 no                | no                           | AT3G44970.1                                                             |                            |              |              |          | no                          | cytochrome p450 superfamily                                                     |                                         |
| PAC-2088731 centromeric |                     |                     | 5                   | 5843745                    | 0.62 no                       | AT3G44060                    |                                                                         |                            |              |              |          | no                          | F-box protein                                                                   |                                         |
| PAC-2088591 centromeric |                     |                     | 5                   | 6507958                    | 0.642347969 no                | no                           | AT3G45252.1                                                             |                            |              |              |          | no                          | gametogenesis related family protein                                            |                                         |
| PAC-20886521            |                     |                     | 5                   | 7711915                    | 0.5332662 no                  | no                           | AT5G62640.2                                                             |                            |              |              |          | no                          | ATELF5 early flowering                                                          |                                         |
| PAC-20886445            |                     |                     | 5                   | 10883265                   | 0.612 no                      | AT3G55677                    |                                                                         |                            |              |              |          | no                          | self-incompatibility protein family                                             |                                         |
| PAC-20887130            |                     |                     | 5                   | 11860286                   | 0.357565402 yes               | no                           | LINE-like TE based on repeat masker                                     | C316257                    |              |              |          | 1228                        | no                                                                              | unknown                                 |
| PAC-20888484            |                     |                     | 5                   | 12269573                   | 0.591900973 yes               | no                           | none                                                                    |                            | 23077        | 11860269     | 11861497 | 3587                        | no                                                                              | Agamous-like 96                         |
| PAC-20910760            |                     |                     | 6                   | 460510                     | 0.481020156 no                | no                           | AT5G11660.1                                                             |                            |              |              |          | no                          | Uclacyanin/basic blue family protein (copper ion binding)                       |                                         |
| PAC-20910395            |                     |                     | 6                   | 1964283                    | 0.525540958 yes               | no                           | AT5G06500                                                               |                            | 18533        | 1963449      | 1973680  | 10231                       | no                                                                              | F box protein                           |
| PAC-20910924            |                     |                     | 6                   | 4662163                    | 0.543298038 no                | no                           | AT5G14345.1                                                             |                            |              |              |          | no                          | alpha/beta hydrolases superfamily                                               |                                         |
| PAC-20911062            |                     |                     | 6                   | 5138537                    | 0.441274937 no                | no                           | AT4G11590.1                                                             |                            |              |              |          | no                          | MEKK protein family                                                             |                                         |
| PAC-20908011            |                     |                     | 6                   | 8230234                    | 0.562011638 no                | no                           | AT5G24210.1                                                             |                            | 10163        |              |          | no                          | unknown                                                                         |                                         |
| PAC-2090851 centromeric |                     |                     | 6                   | 12105457                   | 0.331042412 no                | no                           | AT4G08470.1                                                             |                            |              |              |          | no                          | BP2, involved in chloroplast                                                    |                                         |
| PAC-2090781 centromeric |                     |                     | 6                   | 12310607                   | 0.531285868 no                | no                           | HEUTRON                                                                 |                            |              |              |          | no                          | wound-responsive family protein                                                 |                                         |
| PAC-2090791 centromeric |                     |                     | 6                   | 12756274                   | 0.561780791 no                | no                           | AT2G46080.1                                                             |                            |              |              |          | yes                         | RNc-like cupins superfamily, involved in chloroplast                            |                                         |
| PAC-2090981 centromeric |                     |                     | 6                   | 13055908                   | 0.494412596 yes               | no                           | AT4G10270.1                                                             |                            | 28609        | 13056337     | 13060335 | 3998                        | yes                                                                             | plant invertase/petun methyltransferase |
| PAC-2091101 centromeric |                     |                     | 6                   | 13056631                   | 0.602738051 yes               | AT4G10280                    |                                                                         |                            |              |              |          | no                          | amino acid dehydrogenase family protein, chloroplast                            |                                         |
| PAC-2091031 centromeric |                     |                     | 6                   | 13333813                   | 0.561189346 no                | no                           | AT3G55680.1                                                             |                            |              |              |          | yes                         | APX5, microsomal ascorbate peroxidase                                           |                                         |
| PAC-2090807 centromeric |                     |                     | 6                   | 16377162                   | 0.599648729 no                | AT4G00600                    |                                                                         |                            |              |              |          | yes                         | unknown                                                                         |                                         |
| PAC-20895993            |                     |                     | 7                   | 1610627                    | 0.533394111 no                | AT4G35970                    |                                                                         |                            |              |              |          | yes                         | F-box family protein                                                            |                                         |
| PAC-20895001            |                     |                     | 7                   | 1615155                    | 0.543210773 no                | no                           | AT4G35940.2                                                             |                            |              |              |          | yes                         | unknown                                                                         |                                         |
| PAC-20894314            |                     |                     | 7                   | 1617752                    | 0.504590853 no                | AT4G35930                    |                                                                         |                            |              |              |          | yes                         | F-box family protein                                                            |                                         |
| PAC-2089514 centromeric |                     |                     | 7                   | 13012957                   | 0.307653091 no                | no                           | AT5G32619.1                                                             |                            |              |              |          | no                          | defensin-like protein                                                           |                                         |
| PAC-2089681 centromeric |                     |                     | 7                   | 14138788                   | 0.419877871 no                | no                           | AT5G35330                                                               |                            |              |              |          | no                          | methyl-CpG binding domain containing protein                                    |                                         |
| PAC-20913293            |                     |                     | 8                   | 2859619                    | 0.546735178 no                | no                           | AT1G30790                                                               |                            |              |              |          | no                          | F-box family protein                                                            |                                         |
| PAC-20913904            |                     |                     | 8                   | 6608512                    | 0.61509069 no                 | AT5G50950                    |                                                                         |                            |              |              |          | no                          | fumarate enzyme in mitochondria                                                 |                                         |
| PAC-20911556            |                     |                     | 8                   | 10493204                   | 0.583445 yes                  | AT5G60230                    |                                                                         | scaffold1427356            | 10493065     | 10512708     | 19643    | yes                         | subunit of tRNA splicing endonuclease                                           |                                         |
| PAC-20912733            |                     |                     | 8                   | 10509217                   | 0.6165152 yes                 | no                           | AT5G60250.1                                                             | scaffold1427356            |              |              |          | yes                         | zinc finger family protein                                                      |                                         |

Note: clustered deletions highlighted in yellow clusters with only a single gene shown are due to borderline significant adjacent genes

Table S2 – Enrichment tests of putative inactivating mutations

| Species                  | Mutation set      | GO class | ID         | Term                            | Enrichment | Background | FDR   |
|--------------------------|-------------------|----------|------------|---------------------------------|------------|------------|-------|
| <i>C. bursa-pastoris</i> | Stop codon gained | BP       | GO:0006259 | DNA metabolic process           | 66/1856    | 284/17573  | 9E-05 |
| <i>C. bursa-pastoris</i> | Stop codon gained | BP       | GO:0006310 | DNA recombination               | 18/1856    | 52/17573   | 0.01  |
| <i>C. bursa-pastoris</i> | Stop codon gained | BP       | GO:0006974 | Response to DNA damage stimulus | 36/1856    | 163/17573  | 0.02  |
| <i>C. bursa-pastoris</i> | Stop codon gained | BP       | GO:0022402 | Cell cycle process              | 26/1856    | 103/17573  | 0.03  |
| <i>C. bursa-pastoris</i> | Stop codon gained | BP       | GO:0006508 | Proteolysis                     | 80/1856    | 489/17573  | 0.03  |
| <i>C. bursa-pastoris</i> | Stop codon gained | MF       | GO:0003824 | Catalytic activity              | 817/1885   | 6278/17843 | 2E-09 |
| <i>C. bursa-pastoris</i> | Splice site lost  | BP       | GO:0006259 | DNA metabolic process           | 44/1402    | 284/17573  | 0.03  |
| <i>C. bursa-pastoris</i> | Splice site lost  | MF       | GO:0003824 | Catalytic activity              | 585/1429   | 6278/17843 | 0.001 |
| <i>C. grandiflora</i>    | Stop codon gained | BP       | GO:0006259 | DNA metabolic process           | 33/837     | 284/17573  | 0.001 |
| <i>C. grandiflora</i>    | Stop codon gained | BP       | GO:0006281 | DNA repair                      | 19/837     | 152/17573  | 0.03  |
| <i>C. grandiflora</i>    | Stop codon gained | BP       | GO:0006974 | Response to DNA damage stimulus | 20/837     | 163/17573  | 0.03  |
| <i>C. grandiflora</i>    | Stop codon gained | MF       | GO:0003824 | Catalytic activity              | 391/860    | 6278/17843 | 5E-07 |
| <i>C. grandiflora</i>    | Splice site lost  | MF       | GO:0003824 | Catalytic activity              | 323/749    | 6278/17843 | 0.003 |

Table S3

Geographical origin and collector information for all plant material used in this study.

| Species                  | Sample designation | Locality            | Country     | Collector                      |
|--------------------------|--------------------|---------------------|-------------|--------------------------------|
| <i>C. bursa-pastoris</i> | 70.5               | Artemida            | Greece      | Kate St. Onge                  |
|                          | 5.16               | Valladolid          | Spain       | Santiago Gonzalez-Martinez     |
|                          | 13.16              | Kraków              | Poland      | Sandra Sherwood                |
|                          | 39-12-28           | Bacia               | Italy       | Kate St. Onge                  |
|                          | 12.4               | Halle               | Germany     | Walter Durka                   |
|                          | 16.9               | Nijmegen            | Netherlands | Koen Verhoeven                 |
|                          | VLA                | Vladivostok         | Russia      | Martin Lascoux                 |
|                          | PL                 | Puli, Taiwan        | China       | Y-W Yang                       |
|                          | SE14               | Härnösand           | Sweden      | Svante Holm                    |
|                          | RK32               | Reykjavik           | Iceland     | John Paul Foxe                 |
| <i>C. orientalis</i>     | 1719-3             | Bayan-Olgii Aymag   | Mongolia    | Barbara Neuffer, Herbert Hurka |
|                          | 1719-4             | Bayan-Olgii Aymag   | Mongolia    | Barbara Neuffer, Herbert Hurka |
|                          | 1979-1             | Siberia, Altai Kraj | Russia      | D.A. German, N. Friesen        |
|                          | 1979-7             | Siberia, Altai Kraj | Russia      | D.A. German, N. Friesen        |
|                          | 1981-6             | Siberia, Altai Kraj | Russia      | D.A. German, N. Friesen        |
|                          | 1981-10            | Siberia, Altai Kraj | Russia      | D.A. German, N. Friesen        |
|                          | 1985-1             | Siberia, Altai Kraj | Russia      | D.A. German                    |
|                          | 2008-1             | Xinjiang, Dzungaria | China       | D.A. German et al.             |
|                          | 2008-7             | Xinjiang, Dzungaria | China       | D.A. German et al.             |
|                          | 2008-9             | Xinjiang, Dzungaria | China       | D.A. German et al.             |

Table S4-Genbank accession numbers or PopSet numbers for publicly available sequence data used in this study. Designations are as follows: Cbp A=*C. bursa-pastoris* A homeologs, Cbp B=*C. bursa-pastoris* B homeologs, Cr=*C. rubella*, Cg=*C. grandiflora*, and Np=*N. paniculata*.

| Locus     | Species/homeolog | Popset no./GenBank acc. no.                                            | Reference             |
|-----------|------------------|------------------------------------------------------------------------|-----------------------|
| At1g77120 | Cbp A            | 160334389                                                              | (Slotte et al. 2008)  |
| At1g77120 | Cbp B            | 160334601                                                              | (Slotte et al. 2008)  |
| At1g77120 | Cr               | 160334571                                                              | (Slotte et al. 2008)  |
| At1g77120 | Cg               | 341865754                                                              | (St Onge et al. 2011) |
| At5g10140 | Cbp A            | 160335879                                                              | (Slotte et al. 2008)  |
| At5g10140 | Cbp B            | 160336063                                                              | (Slotte et al. 2008)  |
| At5g10140 | Cr               | 160335193                                                              | (Slotte et al. 2008)  |
| At5g10140 | Cg               | 341866744                                                              | (St Onge et al. 2011) |
| At4g02560 | Cbp A            | 160336247                                                              | (Slotte et al. 2008)  |
| At4g02560 | Cbp B            | 160336429                                                              | (Slotte et al. 2008)  |
| At4g02560 | Cr               | 160335645                                                              | (Slotte et al. 2008)  |
| At4g02560 | Cg               | 341865968                                                              | (St Onge et al. 2011) |
| At4g02560 | Np               | DQ343348.1                                                             | (Slotte et al. 2006)  |
| At4g00650 | Cbp A            | 160335277                                                              | (Slotte et al. 2008)  |
| At4g00650 | Cbp B            | 160335461                                                              | (Slotte et al. 2008)  |
| At4g00650 | Cr               | 160335235                                                              | (Slotte et al. 2008)  |
| At4g00650 | Cg               | FJ650267.1, FJ650266.1, FJ650265.1, FJ650264.1, FJ650263.1, FJ650262.1 | (Guo et al. 2009)     |
| At1g03560 | Cbp A            | 260780309                                                              | (Slotte et al. 2009)  |
| At1g03560 | Cbp B            | 260765968                                                              | (Slotte et al. 2009)  |
| At1g03560 | Cr               | 341606816                                                              | (St Onge et al. 2011) |
| At1g03560 | Cg               | 341605684                                                              | (St Onge et al. 2011) |

|           |      |                                                 |                                           |
|-----------|------|-------------------------------------------------|-------------------------------------------|
| At1g15240 | Cg   | JX065232.1-JX065247.1                           | (St Onge et al. 2012)                     |
|           | Cr   | JQ418746.1-JQ418751.1                           | (St Onge et al. 2012)                     |
|           | Cbp  | JQ418695.1-JQ418745.1                           | (St Onge et al. 2012)                     |
| At2g26730 | Cg   | FJ182827.1-FJ182845.1                           | (Fuxe et al. 2009)                        |
|           | Cr   | FJ182814.1-FJ182826.1                           | (Fuxe et al. 2009)                        |
|           | Cbp  | JQ418752.1-JQ418801.1                           | (St Onge et al. 2012)                     |
| At4g08920 | Cg   | 341606524                                       | (St Onge et al. 2011)                     |
|           | Cr   | 160334783; 341607428                            | (Slotte et al. 2009; St Onge et al. 2011) |
|           | CbpA | 160334825; 260765934                            | (Slotte et al. 2008; Slotte et al. 2009)  |
|           | CbpB | 160335009; 261047491                            | (Slotte et al. 2008; Slotte et al. 2009)  |
| At5g51670 | Cg   | FJ183276.1-FJ183293.1                           | (Fuxe et al. 2009)                        |
|           | Cr   | FJ183262.1-FJ183275.1;<br>JQ418910.1-JQ418914.1 | (Fuxe et al. 2009; St Onge et al. 2011)   |
|           | Cbp  | JQ418859.1-JQ418909.1                           | (St Onge et al. 2012)                     |

Table S5. Demographic parameter estimates with 95% confidence intervals for four models of allopolyploid speciation in *Capsella*. Estimates of effective population sizes for *C. grandiflora* (*Cg*), *C. orientalis* (*Co*), *C. bursa-pastoris* (*Cbp A* and *Cbp B*) are given in thousands of individuals, and estimates of the timing of the origin of *C. bursa-pastoris* (*T1(Cbp)*) and the split between *C. grandiflora* and *C. orientalis* (*T2(Cg-Co)*) are given in kya. Note that for models with exponential population size change, effective population sizes correspond to current effective population sizes.

| Model                             | $N_e Cg$          | $N_e Co$      | $N_e CbpA$    | $N_e CbpB$      | $T1(Cbp)$        | $T2(Cg-Co)$       | $2Nm(Cg-CbpA)$      | $rCg$                                                                        | $rCo$                                                                     | $rCbp$                                                                     |
|-----------------------------------|-------------------|---------------|---------------|-----------------|------------------|-------------------|---------------------|------------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Stepwise change                   | 529<br>(184-692)  | 54<br>(17-67) | 52<br>(18-71) | 104<br>(42-131) | 184<br>(63-251)  | 736<br>(237-981)  | -                   | -                                                                            | -                                                                         | -                                                                          |
| Stepwise change with migration    | 531<br>(230-552)  | 57<br>(24-60) | 6<br>(2-6)    | 102<br>(45-109) | 197<br>(80-211)  | 784<br>(335-847)  | 0.14<br>(0.04-0.52) | -                                                                            | -                                                                         | -                                                                          |
| Exponential change                | 840<br>(148-868)  | 4<br>(1-6)    | 37<br>(6-55)  | 75<br>(12-101)  | 128<br>(22-177)  | 931<br>(161-1159) | -                   | -4.1*10 <sup>-6</sup><br>(-3.5*10 <sup>-5</sup> -<br>-1.3*10 <sup>-6</sup> ) | 2.6*10 <sup>-5</sup><br>(1.6*10 <sup>-5</sup> -<br>1.4*10 <sup>-4</sup> ) | 4.8*10 <sup>-7</sup><br>(-4.0*10 <sup>-5</sup> -<br>9.4*10 <sup>-6</sup> ) |
| Exponential change with migration | 946<br>(383-1077) | 7<br>(4-14)   | 5<br>(3-6)    | 56<br>(53-130)  | 164<br>(100-269) | 914<br>(448-1188) | 0.12<br>(0.06-0.34) | -4.4*10 <sup>-6</sup><br>(-6.2*10 <sup>-6</sup> -<br>-1.3*10 <sup>-6</sup> ) | 1.6*10 <sup>-5</sup><br>(4.3*10 <sup>-6</sup> -<br>2.3*10 <sup>-5</sup> ) | 1.0*10 <sup>-5</sup><br>(-7.3*10 <sup>-6</sup> -<br>6.8*10 <sup>-6</sup> ) |